A PLANAR VERTICAL CHANNEL DMOS STRUCTURE

Richard A. Blanchard

FIELD OF THE INVENTION

This invention relates to a double diffused MOS (DMOS) transistor having a vertical channel region, and in particular to a planar DMOS transistor having a vertical gate.

BACKGROUND OF THE APPLICATION

Double diffused MOS (DMOS) transistors are well known in the prior art. For example, U.S. Patent No. 4,344,081, issued to Pao et al. on August 10, 1982, which is incorporated herein by reference, shows one such prior art structure. Fig. 1 shows a cross section of a prior art N-channel DMOS power transistor. This prior art structure includes an N^-_{3a} epitaxial layer 11 formed on an N^+_{3b} silicon substrate 10. Gate oxide layer 16 is formed on epitaxial layer 11 and doped polysilicon gate 15 is formed on oxide layer 16. Oxide layer 9 covers gate 15. P-type body regions 12a and 12b are diffused into epitaxial layer 11, and N^+_{3c} source regions 13a and 13b are diffused into body regions 12a and 12b, respectively. Source regions 13a and 13b are electrically tied to body regions 12a and 12b by metal contacts 18 and 19, respectively. Contacts 18 and 19 are also electrically tied together. Regions 12c1 and 12c2 beneath gate 15 in body regions 12a and 12b, respectively, are channel regions. When the potential between gate 15 and source regions 13a and 13b is sufficiently high and with a positive voltage on drain contact 17, carriers flow laterally from source regions 13a and 13b through channel regions 12a and 12b, respectively, to drain region 11 and then vertically downward through drain region 11 and N^+_{3c} substrate 10 to drain contact 17, as indicated by arrows 20a and 20b in Fig. 1. P-channel DMOS transistors have a similar structure, but P-type and N-type regions are reversed, and a voltage of the opposite sign produces current flow.

As explained above, the carriers that flow in the prior art vertical DMOS transistors shown in Fig. 1 must change direction, first flowing laterally and then vertically. Carrier flow is more efficient if the source, body and drain regions are arranged vertically as shown in Fig. 2. Fig. 2 shows a cross section of a prior art DMOS transistor with a U shaped gate extending into the epitaxial layer. This structure is due to Ueda et al. and is explained in more detail in A New Vertical Power MOSFET Structure with Extremely Reduced On-Resistance, IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. ED-32, NO. 1, January 1985, which is incorporated herein by reference. In this prior art structure N^- epitaxial layer 11 is again formed on N^+ substrate 10. A P-type dopant is diffused into epitaxial layer 11 and an N^+-type dopant is diffused into a portion of the epitaxial layer that has been doped with a P-type dopant. Rectangular groove 23, having vertical walls, is then etched in the epitaxial layer using reactive ion beam etching, thereby creating P-type body regions 20a and 20b and corresponding N^+ source regions 21a and 21b as shown in Fig. 2. Source regions 21a and 21b are electrically tied to body regions 20a and 20b, respectively, by metal contacts 18 and 19 which are also electrically tied together. A slight wet etch is then applied to smooth the surface of groove 23. Gate oxide 24 is formed in rectangular groove 23, and a U shaped polysilicon gate 25 is formed over gate oxide 24.

The prior art structure of Fig. 2 has the advantage that when the gate to source potential is sufficient to turn on the transistor, carriers flow vertically from N^+ source regions 21a and 21b through channel regions 22c1 and 22c2 in body regions 20a and 20b, respectively, and continue to flow vertically downward through drain region 11 to N^+ substrate 10 and drain contact 17. However, the structure of Fig. 2 has a disadvantage in that it is difficult to fabricate because it requires the formation of a U-shaped gate and results in a transistor with a nonplanar surface.

SUMMARY OF THE INVENTION

A DMOS device is disclosed which has a vertical gate and a planar surface. The device has a flat surface for all masking steps while still allowing contact to be made to the vertical gate.

In one embodiment, a DMOS power transistor is disclosed which has a drain region of a first conductivity type, a body region of a second conductivity type formed above the drain region, and a source region of first conductivity type. An upward opening rectangular groove extends downward through the source and body regions and into the drain region so that a first source region in a first body region lies on one side of the rectangular groove and a second source region in a second body region lies on the other side of the rectangular groove.

The upward opening rectangular groove is lined with an upward opening dielectric region which is filled with the gate region so that a vertical gate is formed having a top surface which lies between the first and second source regions. An insulating layer is then formed over the above structure so that a transistor with a planar surface is obtained. In operation, carriers flow vertically between the source and drain regions.

In another embodiment, a semiconductor device having a vertical gate region is formed in a block of semiconductor material. The vertical gate region lies in an upward opening dielectric region which lines an upward opening rectangular groove. A source region of a first conductivity type is formed above a body region of a second conductivity type which lies above a drain region of a first conductivity type. The source, body and drain regions are all adjacent one vertical surface of the dielectric material. The top surface of the vertical gate region lies opposite the source region and the bottom surface of the gate region lies opposite the drain region. The second embodiment also includes an insulating layer formed over the gate, source and body regions resulting in a device with a planar top surface. Still other embodiments are described below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a prior art N-channel DMOS transistor;

Figure 2 shows a prior art DMOS transistor having vertical channel regions;

Figure 3 shows one embodiment of the DMOS structure of the present invention;

Figures 4a through 4f show process steps in the formation of the transistor shown in Fig. 3;

Figure 5 shows a second embodiment of the DMOS transistor of the present invention;

Figure 6 shows a third embodiment of the DMOS transistor of the present invention;

Figure 7 shows a cross-section of an insulated gate transistor formed according to the present invention;

Figure 8 shows a cross-section of an MOS-gated silicon controlled rectifier formed according to the present invention; and

Figure 9 shows a top view of one surface geometry employed by the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Figure 3 shows one embodiment of the vertical gate planar DMOS power transistor of the present invention. The process sequence for fabricating this N-channel transistor is shown in Figs. 4a-4f. In other embodiments, the vertical gate planar DMOS transistor of the present invention is a P-channel device.

Substrate 10 shown in Fig. 4a is a silicon wafer doped with N-type impurities so that its resistivity is within the range of 0.005 to 0.1 ohm-cm (in one embodiment, 0.02 ohm-cm). An N-type epitaxial layer 11 having a resistivity between 0.2 and 100 ohm-cm (in one embodiment, 2.4 ohm-cm) is then grown on substrate 10 to a thickness between 6 and 150 microns (in one embodiment, 13.5μ). Substrate 10 and epitaxial layer 11 typically have a [100] crystal orientation.

A layer of dielectric material (not shown) is then formed over the wafer by heating the wafer in an oxygen

atmosphere at about 900-1200°C to form a silicon dioxide layer approximately 1,000-10,000Å thick over the surface of the wafer. Using standard photoresist techniques, a body mask pattern is transferred to the surface of the silicon dioxide layer which exposes those regions of the silicon dioxide layer through which ions are implanted to form P region 20. (In one version, no body mask is needed, and P-type dopant is implanted to form region 20.) In one embodiment, P region 20 is formed by implanting boron ions at a dosage between 10^{13} and 2×10^{14} ions/cm^2 at an energy level between 40 and 120 KEV and then annealing the structure for approximately 4 to 12 hours in an atmosphere of oxygen or nitrogen. P region 20 typically ranges in depth from 2 to 4 microns. Alternatively, P region 20 may be formed using standard diffusion techniques.

After a source mask is applied to the wafer, the wafer may then be etched to thin or remove the silicon dioxide formed during the annealing process. (If the oxide formed after the body diffusion is not too thick, i.e., it does not block the implant, no thinning etch is needed.) An N-type ion implantation is performed using arsenic or phosphorus ions at a dosage between 5×10^{14} and 1×10^{16} ions/cm^2 using an implant energy of 50 to 150 KEV. The wafer is then annealed at a temperature of 900-1200°C for approximately 0.5 to 3 hours in an atmosphere of oxygen and nitrogen to form N^+ source region 21 which typically ranges in depth from 1 to 2 microns.

The silicon dioxide layer 30 is formed during the above source drive-in. A gate groove mask (not shown) is applied, and the SiO_2 layer is etched using buffered HF.

As shown in Fig. 4b, rectangular groove 31 having vertical sidewalls is then etched using reactive ion beam etching or other etching technique which permits anisotropic etching regardless of crystallographic orientation. Groove 31 preferably has a width less than or equal to 2µm and a depth between 3µm and 10µm. As shown in Fig. 4c, gate dielectric layer 32 is then formed over the surfaces of the groove 31. In one embodiment, gate dielectric layer 32 is

silicon dioxide having a thickness in a range of 500~1000Å and is formed by heating the wafer in an oxygen ambient containing water at a temperature of 900~1100°C for 0.5 to two hours (in another embodiment, insulating layer 32 is a combination of silicon dioxide and silicon nitride which is either grown or deposited). The gate dielectric forms an inner, upward opening, rectangular groove 31*.

Polysilicon layer 33 (shown in Fig. 4d) is then deposited using a low pressure chemical vapor deposition process (LPCVD) to a thickness sufficient to fill rectangular groove 31*. For example, if groove 31 is 1.5 microns wide and 6 microns deep, polycrystalline silicon layer 33 is deposited with a thickness of 1~2µm. Polycrystalline silicon layer 33 is doped either during deposition or subsequent to deposition, typically using phosphorus, so that it has a sheet resistance of between 30 and 50 ohms/square. Alternatively, in another embodiment, layer 33 comprises a layer of silicide formed using conventional techniques to a depth sufficient to fill rectangular groove 31*.

Polycrystalline silicon layer 33 is then subjected to a CF_4 etch or another etch technique without using a mask in the trenched area, groove 31*, (except for a mask (not shown) which may be placed at any convenient point along the length of groove 31 in order to keep a contact pad (not shown) to the to-be-formed gate 34 shown in Fig. 4e) in order to remove the polycrystalline silicon not in groove 31*. The portion of polycrystalline silicon layer 33 remaining in groove 31* after the CF_4 etch is denoted by 34 in Fig. 4e and serves as the gate of the vertical DMOS transistor. The etch is continued until top surface 34a lies 0.25~0.5µm below the top surface of layer 30. This top surface depth is controlled by etch time past clearing the field. The etch must be terminated so that gate 34 overlaps N^+ regions 21a and 21b shown in Fig. 4f after the subsequent oxidizing step. The wafer is then oxidized in an atmosphere containing oxygen (which consumes a portion of polysilicon layer 33 in groove 31*) until the top surface of the

oxidized portion 35 above gate 34 forms an essentially flat (planar) surface with the top surface of passivating layer 30 whose thickness may also be slightly increased during the formation of region 35. Of importance, the etch to form surface 34a must be terminated sufficiently soon so that after the oxidation which forms silicon dioxide layer 35, the top portion of gate 34 overlaps N^+_{30} source regions 21a and 21b (see Fig. 4f).

The above structure has a flat surface for all masking steps while still allowing contact to be made to the gate region. The source/body contact shown schematically in Fig. 3 is fabricated using prior art techniques, and in cross section typically appears as shown in Fig. 1.

When the gate-to-source potential is sufficiently high and with a positive potential on drain 17 (Fig. 3), electrons flow vertically from N^+_{30} source regions 21a and 21b through channel regions 22c1 and 22c2 in body regions 20a and 20b, respectively, and continue to flow vertically downward through drain 11 and N^+_{30} substrate (drain) 10 to drain contact 17.

Typically, many DMOS devices similar to the one shown in cross section in Fig. 3 are formed simultaneously. Layout efficiency varies with surface geometry. There is a wide variety of layouts. Fig. 9 shows a top view of one surface geometry employed by this invention, namely, a square source and body region on a square gate grid 35. In Fig. 9, S denotes the locations of the source regions, B the locations of the body regions, and G the locations of the gate regions. The dotted line shown in Fig. 9 corresponds to the cross section shown in Fig. 3. In another layout, (not shown) the gate and source and body regions are interdigited. Another layout (not shown) has hexagonal source and body regions on an hexagonal gate grid. Still another layout employs square source and body regions on a hexagonal gate grid. The latter layout is more efficient than the others. Other source geometries include rectangles, circles and triangles.

The structure shown in Fig. 3 reduces the total area

requirement from 30% to 50% below that of the Ueda device shown in Fig. 2.

Fig. 5 shows a cross section of an alternate embodiment of the invention in which a P^-_{31} epitaxial layer 40 is formed on substrate 10 in place of N^-_{31} epitaxial layer 11 and in which gate 34 in groove 31 reaches through to the N^+_{30} substrate. In this embodiment the P^-_{31} epitaxial layer serves as the body region of the transistor. This embodiment results in a transistor having a lower on resistance than the device shown in Fig. 4a by as much as a factor of 2 and a lower breakdown voltage (typically around 30 volts) than the device shown in Fig. 4a.

Fig. 6 shows another alternate embodiment of the invention which is similar to the embodiment shown in Fig. 4f except that groove 31 is formed sufficiently deep so that gate 34 extends completely through the epitaxial layer 11 and into substrate region 10. This alternate embodiment also has a lower on resistance and a lower breakdown voltage (typically about 30 volts) than the device shown in Fig. 4a. This is acceptable in low voltage applications, for example, low voltage motors and Schottky diode replacements.

The above description has been given in terms of DMOS transistors, but the invention also applies to other MOS-gated devices such as an MOS-gated SCR or a MOS-gated conductivity modulated device.

Fig. 7 shows a cross section of one embodiment for an insulated gate transistor. Fig. 8 shows a cross section of one embodiment for an MOS-gated silicon controlled rectifier.

The silicon controlled rectifier shown in Fig. 8 is fabricated in the same manner as explained above for the transistor shown in Fig. 3 in connection with Figs. 4a through 4f except that the starting material is a silicon substrate 41 heavily doped with P-type material, for example doped with Boron to a resistivity of 0.01 ohm-cm.

In operation, silicon controlled rectifier 60 is switched on by appropriately biasing source/body terminal 50 (shown schematically in Fig. 8), which provides electrical

contact to source regions 21a and 21b and body regions 20a and 20b, gate terminal 49, which contacts gate 34, and substrate contact 51. Gate contact 49 and substrate contact 51 are biased positive relative to source/body contact 50. MOSFET mode conduction is initiated by electrons flowing from source regions 21a and 21b through channel regions 22c1 and 22c2, respectively, to N^-_{31} drain region 11.

Drain region 11 also serves as the base for the PNP bipolar (junction) transistors comprising emitter region 20a, base 11, and collector region 41; and emitter region 20b, base 11, and collector region 41, respectively. With substrate contact 51 biased positively with respect to source/body contact 50 and with electrons flowing into base 11, the bipolar transistors are triggered into conduction and may latch in the on state even when the gate bias is removed.

The insulated gate transistor shown in Fig. 7 is similar to the structure shown in Fig. 8 except that in order to suppress thyristor action, the IGT shown in Fig. 7 is designed with narrow N^+_{30} source regions 21a and 21b which reduce the lateral body resistance beneath the source regions. See The Insulated Gate Transistor: A New Three-Terminal MOS-Controlled Bipolar Power Device, IEEE Trans. on Electron Devices, Vol. ED-31 No. 6, June 1984, which is incorporated herein by reference. For example, the width w'_{40} of N^+_{30} source regions 21a and 21b in Fig. 7 is typically between 2μm and 4μm whereas the width w of N^+_{30} source regions 21a and 21b in Fig. 8 is typically between 6μm and 8μm. The doping profile of the body regions beneath source regions 21a and 21b in Fig. 7 is also selected to reduce lateral body resistance beneath source regions 21a and 21b. See Blanchard, U.S. Patent No. 4,345,265, issued August 17, 1982, which is incorporated herein by reference. A low lateral body resistance beneath source regions 21a and 21b in Fig. 7 prevents the NPN transistor formed by source regions 21a and 21b, body regions 22c1 and 22c2, and the N-type drain region 11 from becoming active. As long as this NPN bipolar transistor does not turn on, the

regenerative action characteristic of an SCR does not occur.

The above embodiments are meant to be exemplary and not limiting. In view of the above disclosure, many modifications and substitutions will be obvious to one of average skill in the art without departing from the scope of the invention.

What I claim;